Markov processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(E, \mathcal{E})$ be a measurable space.

In these notes $(E, \mathcal{E})$ will be either

1. E a discrete set (namely finite or countable such as a subset of $\mathbb{N}^d$, $\mathbb{Z}^d$) endowed with the σ-algebra $\mathcal{E}$ of all subsets,

2. $\mathbb{R}^d$ endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$.

A *stochastic process* $(X_t)_t$, with values in E, is a collection of E-valued random variables indexed by a time index t.

The time index t belongs to some totally ordered set such as a subset of $\mathbb{Z}$ or $\mathbb{R}$.

Definition 1 *The stochastic process $(X_t)_t$ is a* Markov chain *if, for all $m \geq 1$ and times $t_1 < t_2 < \cdots < t_{m+1}$ and subsets $E_1, E_2, \ldots, E_{m+1} \in \mathcal{E}$ one has*

$$\begin{aligned} &\mathbb{P}\{ X_{t_{m+1}} \in E_{m+1} \mid X_{t_m} \in E_m, \ldots, X_{t_1} \in E_1 \} \\ &= \mathbb{P}\{ X_{t_{m+1}} \in E_{m+1} \mid X_{t_m} \in E_m \} \end{aligned} \tag{1}$$

The Markov chain $(X_t)_{t\geq 0}$ is called time homogeneous *if transition probabilities*

$$\mathbb{P}\{ X_{t+s} \in E_2 \mid X_s \in E_1 \}$$

do not depend on s.

Note that Markov property simply means conditional independence of the future with respect to the past. Information on the process today depends on the past history only through the 'last time' in the past and not on the whole past history.

In the case where the time is discrete, i.e. it can be written explicitly as $\{t_0, t_1, t_2 \ldots\}$, and the set E is at most countable we recover our first definition of a discrete time Markov chain (consider $\mathcal{E}$ the σ-algebra of all subsets of E and sets $E_m = \{e_m\}$, $e_m \in \mathcal{E}$).

In the general case we introduce *transition kernels*, a useful tool for writing transition probabilities of Markov chains on a continuous state space $(E, \mathcal{E})$.

Definition 2 *Let $(X_t)_{t\geq 0}$ be Markov chain with values in a (continuous) state space E. Transition kernels are functions $P_t : E \times \mathcal{E} \to [0,1]$ such that*

1. *$P_t(x, \cdot)$ is a probability measure for all $x \in E$,*
2. *$P_t(\cdot, A)$ is a measurable function for all $A \in \mathcal{E}$,*
3. *the conditional distribution of X_{t+s} given X_s is given by*

$$\mathbb{P}\{\, X_{t+s} \in A \mid X_s = x \,\} = \int_A P_t(x, dy) \tag{2}$$

Remark. In the case of a continuous time Markov chain with discrete state space item 3 in the above definition reads as

$$p_{ij}(t) = \mathbb{P}\{\, X_{t+s} = j \mid X_s = i \,\} = \int_{\{j\}} P_t(i, dy) = P_t(i, \{j\})$$

Remark. Equation (2) can be also expressed as

$$\mathbb{E}[\, X_{t+s} \mid \sigma(X_s) \,] = \int_E y\, P_t(X_s, dy)$$

Remark. Clearly, for all $r < s < t$

$$\mathbb{P}\{\, X_t \in A \mid X_r = x \,\} = P_{t-r}(x, A)$$

and also

$$\begin{aligned} \mathbb{P}\{\, X_t \in A \mid X_r = x \,\} &= \int_E \mathbb{P}\{\, X_t \in A \mid X_s = y \,\} P_{s-r}(x, dy) \\ &= \int_A P_{t-s}(y, A) P_{s-r}(x, dy) \end{aligned}$$

Therefore we get the Chapman-Kolmogorov equation for kernels

$$P_{t-r}(x, A) = \int_A P_{t-s}(y, A) P_{s-r}(x, dy)$$

and, in the discrete time case,

$$P_{n+m}(x, A) = \int_A P_n(y, A) P_m(x, dy).$$

Example 1 *(First-order autoregressive Gaussian process)* A widely used model in social sciences is the following process

$$X_{n+1} = b\, X_n + \sigma Z_{n+1} \qquad X_0 = x$$

where $x_0, b \in \mathbb{R}$ and $(Z_n)_{n\geq 1}$ are independent identically distributed standard Gaussian random variables.

This defines a Markov chain $(X_n)_{n\geq 0}$ with state space $\mathbb{R}$ (considered as a mesurable space with the Borel σ-algebra) and, since the conditional distribution of X_{n+1} given X_n is a Gaussian distribution $N(bX_n, \sigma^2)$, the *one step* transition kernel

$$\begin{aligned} P(x, A) &= \mathbb{P}\{ X_{n+1} \in A \mid X_n = x \} \\ &= \int_A \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{y - bx}{2\sigma^2}\right) dy. \end{aligned}$$

Example 2 *(Brownian motion)* Let $(B_t)_{t\geq 0}$ be a family of random variables with $B_0 = 0$ and the following properties:

1. For all n and all $0 < t_1 < t_2 < \cdots < t_n$ random variables (increments)

   $$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

   are independent,

2. For all $s < t$ the random variable $B_t - B_s$ has Gaussian distribution $N(0, \sigma^2(t-s))$.

Using the independence of increments one immediately checks that this is a time homogeneous Markov process with transition kernels

$$P_t(x, A) = \mathbb{P}\{ B_{t+s} \in A \mid B_s = x \} = \frac{1}{\sqrt{2\pi t}\,\sigma} \int_A \mathrm{e}^{-\frac{(y-B_s)^2}{2\sigma^2 t}} dy$$

1 Stationary distributions

Given a Markov process $(X_t)_t$ with transition kernels P_t, if we denote by μ_t the distribution of X_t, then , for all s, t and all bounded measurable functions f, g

$$\mathbb{E}\left[f(X_{t+s})g(X_s)\right] = \int_E g(x)\mu_s(dx) \int_E f(y)P_t(x, dy)$$

In particular, if we consider $f = 1_A$ and g a constant function, we find

$$\mu_{t+s}(A) = \mathbb{P}\{ X_{t+s} \in A \} = \int_E \mu_s(dx)P_t(x, A).$$

The above identity shows the relationship between the distribution of X_s and X_{t+s} and motivates the following

Definition 3 *Let $(X_t)_t$ be a time homogeneous transition Markov process with transition kernels $(P_t)_t$.*

- *A σ-finite measure μ on $\mathcal{E}$ is said to be invariant with respect to the transition kernels $(P_t)_t$ if*

$$\mu(A) = \int_E P_t(x, A)\mu(dx), \qquad \forall A \in \mathcal{E}$$

- *If μ is a probability measure, then it is said to be an invariant (or stationary) distribution for the Markov process $(X_t)_t$.*

Example 3 *(Autoregressive Gaussian process)* Consider the Markov process introduced in Example 1 but suppose that also X_0 is a random variable. If $|b| < 1$ it is natural to conjecture that, for n big enough, the distribution of X_n (and X_{n+1}) will be approximately Gaussian with zero mean. Therefore a Gaussian distribution $N(0, \theta)$ is a natural candidate. Now, if two independent random variables Y, Z satisfy $Y = bY + \sigma Z$, then the characteristic function $\phi_Y(t) = \mathrm{e}^{-\theta^2 t^2/2}$ of Y satisfies

$$\mathrm{e}^{-\theta^2 t^2/2} = \mathrm{e}^{-b^2\theta^2 t^2/2}\mathrm{e}^{-\sigma^2 t^2/2}.$$

It follows that $\theta^2 = \sigma^2/(1-b^2)$. It is now straightforward to check that, for all b with $|b| \neq 1$, if $X_n \sim N(0, \sigma^2/(1-b^2))$ and Z_{n+1} is independent of X_n, then $X_{n+1} = bX_n + \sigma Z_{n+1}$ has also Gaussian distribution $N(0, \sigma^2/(1-b^2))$ for all n. This shows that the Gaussian distribution $N(0, \sigma^2/(1-b^2))$ is an invariant distribution.

Example 4 *(Brownian motion)* The Brownian motion $(B_t)_{t\geq 0}$ with $B_0 = 0$ clearly has no invariant distribution because all random variables B_t are Gaussian but with different variances. It would not be difficult to show that, also if we consider another initial condition by putting $X_t = X_0 + B_t$ for some random variable X_0 independent of all B_t, then the Markov process $(X_t)_{t\geq 0}$ has no invariant distribution.

2 Irreducibility

For Markov chains on a continuous state space irreducibility is defined with respect to a reference measure.

Definition 4 *Let φ be a measure on $\mathcal{E}$. A Markov chain $(X_t)_t$ on a state space E with transition kernels $P_t(\cdot, \cdot)$ is called φ-irreducible if, for all $A \in \mathcal{E}$ with $\varphi(A) > 0$ there exists an $t > 0$ such that $P_t(x, A) > 0$ for all $x \in E$.*

Remark. Let ϕ be any measure equivalent to the Lebesgue measure on $\mathbb{R}$, namely a measure φ which is absolutely continuous with respect to the Lebesgue measure such that also the Lebesgue measure is absolutely continuous with respect to φ (or, in other words, such that a Borel set A satisfies

$\varphi(A) = 0$ if and only if A has Lebesgue measure equal to 0. Both previous Markov chains are ϕ-irreducible because $P_t(x, A) > 0$ for all $x \in E$, $t > 0$ and all Borel subset A of $\mathbb{R}$ with $\varphi(A) > 0$.

Remark. If Markov chain is irreducible with respect to several measures one can find another measure φ^* such that the Markov chain is φ^*-irreducible and all the other irreducibility measures are absolutely continuous with respect to φ^*.

3 Recurrence

For Markov processes on a continuous state space also recurrence is defined with respect to a reference measure. We consider, for simplicity, only the case of a discrete time Markov chains $(X_n)_{n\geq 0}$.

For all $A \in \mathcal{E}$ we consider the random variable

$$N_A = \sum_{n\geq 0} \mathbf{1}_A(X_n) = \sum_{n\geq 0} \mathbf{1}_{\{X_n \in A\}}$$

number of visits in A.

Definition 5 *For a Markov process $(X_n)_{n\geq 0}$*

- *Suppose that $X_0 = x_0 \in A$. A set $A \in \mathcal{E}$ is called* Harris recurrent *if $\mathbb{P}_x\{\, N_A = +\infty \,\} = 1$ for all $x \in A$,*
- *The Markov chain $(X_n)_{n\geq 0}$ is* Harris recurrent *if there exists a measure φ such that it is φ irreducible and all subsets A of E with $\varphi(A) > 0$ are Harris recurrent.*

The condition $\mathbb{P}_x\{\, N_A = +\infty \,\} = 1$ (infinite number of visits in A), in many situations, is equivalent to $\mathbb{E}\,[N_A] = +\infty$ (*mean* infinite number of visits in A) as for discrete state and discrete time Markov chains.

The following is a sufficient condition for Harris recurrence

Proposition 6 *Consider a Markov process $(X_t)_t$ with state space E and transition kernels $(P_t(\cdot,\cdot))_t$. If*

1. *the chain is φ-irreducible, for a probability measure φ,*
2. *φ is the unique, invariant distribution of the Markov process,*
3. *the measure $P_t(x,\cdot)$ is absolutely continuous with respect to φ for all $x \in E$,*

then the Markov process $(X_t)_t$ is Harris recurrent.

It can be shown that a φ-irreducible Harris recurrent Markov process with a unique stationary probability distribution converges in distribution. We illustrate it by an example.

Example 5 *(Autoregressive Gaussian process)* Let $x_0 \in \mathbb{R}$, $b \in \mathbb{R}$ with $|b| < 1$, $\sigma > 0$ and define inductively

$$X_{n+1} = b\,X_n + \sigma Z_{n+1} \qquad X_0 = x$$

where $\sigma > 0$ and $(Z_n)_{n\geq 1}$ are independent identically distributed standard Gaussian random variables. Note that

$$\begin{aligned} X_n &= bX_{n-1} + \sigma Z_n \\ &= b^2 X_{n-2} + b\sigma Z_{n-1} + \sigma Z_n \\ &= b^3 X_{n-3} + b^2\sigma Z_{n-2} + b\sigma Z_{n-1} + \sigma Z_n \\ &= \cdots \\ &= b^n x_0 + \sigma \sum_{k=1}^{n} b^{n-k} Z_k. \end{aligned}$$

The characteristic function of X_n is

$$\begin{aligned} \phi_{X_n}(t) &= \exp(\mathrm{i}b^n x_0 t) \prod_{k=1}^{n} \phi_{Z_k}(b^{n-k}\sigma t) \\ &= \exp(\mathrm{i}b^n x_0 t) \exp\left(\frac{\sigma^2 t}{2} \sum_{k=1}^{n} b^{2(n-k)}\right) \\ &= \exp(\mathrm{i}b^n x_0 t) \exp\left(\frac{\sigma^2 t}{2} \cdot \frac{1-b^{2n}}{1-b^2}\right) \end{aligned}$$

It follows that (for $|b| < 1$)

$$\lim_{n\to\infty} \phi_{X_n}(t) = \exp\left(\frac{\sigma^2 t}{2(1-b^2)}\right)$$

and $(X_n)_{n\geq 0}$ converges in law to $N(0, \sigma^2/(1-b^2))$.

Remark. It does not converge in L^2 since

$$X_n - X_m = b^n x_0 - b^m x_0 + \sigma \sum_{k=1}^{m} (b^{n-k} - b^{m-k}) Z_k + \sigma \sum_{k=m+1}^{n} b^{n-k} Z_k$$

Therefore

$$
\begin{aligned}
\mathbb{E}\left[\,|X_n - X_m|^2\,\right] &= |b^n - b^m|^2\,|x_0|^2 + \sigma^2 \sum_{k=1}^{m} \left|b^{n-k} - b^{m-k}\right|^2 + \sigma^2 \sum_{k=m+1}^{n} b^{2(n-k)} \\
&= |b^n - b^m|^2\,|x_0|^2 + \sigma^2\,|b^n - b^m|^2\,\frac{b^{-2(m+1)} - b^{-2}}{b^{-2} - 1} \\
&+ \sigma^2 \sum_{h=0}^{n-m-1} b^{2h} \\
&= |b^n - b^m|^2\,|x_0|^2 + \sigma^2 \frac{|b^{n-m-1} - b^{-1}|^2 - |b^{n-1} - b^{m-1}|^2}{b^{-2} - 1} \\
&+ \sigma^2 \frac{1 - b^{2(n-m)}}{1 - b^2}
\end{aligned}
$$

and

$$
\lim_{n,m\to\infty} \mathbb{E}\left[\,|X_n - X_m|^2\,\right] = \sigma^2 \left(\frac{b^{-2}}{b^{-2}-1} + \frac{1}{1-b^2}\right) = \frac{2\sigma^2}{1-b^2} > 0.
$$

4 Law of large number for Markov processes

Theorem 7 *Let $(X_n)_{n\geq 0}$ be a Harris recurrent Markov chain with invariant measure φ. Consider a measurable function f such that*

$$
\int_E |f(x)|\,\varphi(dx) < \infty.
$$

Then

$$
\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \int_E f(x)\varphi(dx)
$$

almost surely.